

A new method for obtaining explicit estimators in unbalanced mixed linear models

Tatjana von Rosen¹ · Dietrich von Rosen^{2,3}  · Julia Volaufova⁴

Received: 17 September 2016 / Revised: 3 July 2017
© The Author(s) 2017. This article is an open access publication

Abstract The general unbalanced mixed linear model with two variance components is considered. Through resampling it is demonstrated how the fixed effects can be estimated explicitly. It is shown that the obtained nonlinear estimator is unbiased and its variance is also derived. A condition is given when the proposed estimator is recommended instead of the ordinary least squares estimator.

Keywords Linear mixed models · Explicit estimators · Ordinary least squares estimators · Maximum likelihood estimators · Abstract bootstrapping

1 Introduction

Unbalanced data have challenged statisticians for decades. In a simple two-way analysis of variance model with interactions and an unequal number of replicates per cell, which is often referred to as the unbalanced case, it is not obvious at all how to test

✉ Dietrich von Rosen
dietrich.von.rosen@slu.se

Tatjana von Rosen
Tatjana.vonRosen@stat.su.se

Julia Volaufova
jvolau@lsuhsc.edu

¹ Department of Statistics, Stockholm University, Stockholm, Sweden

² Department of Engineering and Technology, Swedish University of Agricultural Sciences, Uppsala, Sweden

³ Department of Mathematics, Linköping University, Linköping, Sweden

⁴ Biostatistics Program, LSU Health - New Orleans, School of Public Health, New Orleans, USA

various hypotheses and estimate effects. For this model, however, it is always possible to construct F-distributed test statistics, under standard normality assumptions, although there remains a serious issue of interpreting the results. Approximate tests in non-normal unbalanced two-way random models has been studied by Güven (2012).

In case of a two-way analysis of variance model with random effects the situation starts to become even more challenging since it is not obvious how to set up the test statistic so that inference can rely on the F-distribution. In this context, the idea to sacrifice some power in order to create a balanced data set proposed by Khuri (1986), Khuri and Littell (1987), Khuri (1990), Gallo and Khuri (1990), Öfversten (1993, 1995) and Christensen (1996) is very appealing.

The above mentioned works were mainly concerned with creating tests for variance components and estimators of fixed effects in the two-way model with interactions with the key idea that one can resample from the residuals to increase variation in order to mimic a balanced model.

In this article we adopt the above mentioned ideas and focus on estimation of fixed effects parameter in a mixed linear model with two variance components. The new proposed estimator can be considered as an alternative to the classical moment estimator, e.g., see Henderson (1953) or Al Sarraj and von Rosen (2009), where in the latter reference improved moment estimators were discussed although the focus was on variance components estimation.

Throughout the article, vectors (matrices) will be denoted by bold letters (bold upper cases), $\mathcal{C}(\cdot)$ denotes the column vector space, and for any matrix \mathbf{A} , $P_{\mathbf{A}}$ denotes the orthogonal projection onto $\mathcal{C}(\mathbf{A})$. $\sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ stands for distributed as a p -dimensional multivariate normal distribution with mean $\boldsymbol{\mu}$ and dispersion matrix $\boldsymbol{\Sigma}$. For a random vector (variable) $E[\cdot]$ stands for its expectation, $D[\cdot]$ for the dispersion matrix (variance), and $\text{cov}[\cdot, \cdot]$ for the covariance between two vectors (random variables). Other notation will be defined in the subsequent text as needed.

2 Model

The main focus in this article is on the explicit estimation of $\mathbf{X}\boldsymbol{\beta}$, the vector of fixed effects parameters, in the following mixed linear model,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon}, \quad (2.1)$$

where \mathbf{y} is an $n \times 1$ observable random vector. Let \mathbf{X} be an $n \times p$, known model matrix of rank $\text{rank}(\mathbf{X}) = k \leq p < n$ and \mathbf{Z} an $n \times m$, $m < n$ known matrix, such that $\mathcal{C}(\mathbf{Z}) \not\subset \mathcal{C}(\mathbf{X})$.

Here, $\boldsymbol{\beta} \in R^p$ is unknown and fixed parameter of the mean. Furthermore, the vector of random effects, $\boldsymbol{\gamma} \sim N_m(\mathbf{0}, \sigma_{\gamma}^2 \mathbf{I}_m)$, is assumed to be independently distributed of the random errors $\boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$; the variances $\sigma_{\gamma}^2 \geq 0$ and $\sigma^2 > 0$ are unknown scalar parameters. Note that the model comprises the important case when $E[\boldsymbol{\gamma}] \neq \mathbf{0}$ because the mean of $\mathbf{Z}\boldsymbol{\gamma}$ can be “moved into” $\mathbf{X}\boldsymbol{\beta}$.

Let $\rho = \sigma_{\gamma}^2 / \sigma^2$. The covariance matrix of \mathbf{y} is then $D[\mathbf{y}] = \sigma^2(\rho \mathbf{Z}\mathbf{Z}' + \mathbf{I}_n)$. The ordinary least squares estimator (OLS) is always available for $\mathbf{X}\boldsymbol{\beta}$, i.e., $\widetilde{\mathbf{X}}\boldsymbol{\beta} = P_{\mathbf{X}}\mathbf{y}$,

but it is known from maximum likelihood theory that in general, the weighted least squares estimator asymptotically performs better. In model (2.1), explicit closed form maximum likelihood estimators are available only in special cases. Therefore, in this article the aim is to find an explicit estimator for $X\beta$, which also uses the random effects variation due to the presence of random effects γ .

3 Main result

Denote $l = \text{rank}(X : Z)$, then the dimension of the column space $\mathcal{C}(X)^\perp \cap \mathcal{C}(X : Z)$ is $l - k$. Let A be an $n \times (l - k)$ matrix, such that the column vector space satisfies $\mathcal{C}(A) = \mathcal{C}(X)^\perp \cap \mathcal{C}(X : Z) = \mathcal{C}(P_{(X:Z)} - P_X)$ and $A'A = I_{l-k}$. The matrix A can be obtained through the Gram-Schmidt orthogonalization algorithm.

Denote B_1 an $n \times k$ matrix, such that $B_1B_1' = P_X$ and $B_1'B_1 = I_k$. Let $B_2 = A(A'ZZ'A)^{-1/2}$, and let B_3 be any $n \times (n - l)$ matrix, such that $\mathcal{C}(B_3) = \mathcal{C}(X : Z)^\perp$ with $B_3'B_3 = I_{n-l}$. From the assumptions it follows that $A'ZZ'A$ is positive definite and therefore its inverse exists: $\text{rank}(A'ZZ'A) = \text{rank}(A'Z) = \text{rank}(A : Z^o) - \text{rank}(Z^o)$, where Z^o is any matrix spanning $\mathcal{C}(Z)^\perp$; since A is of full column rank, the statement is true because $\mathcal{C}(A) \cap \mathcal{C}(Z)^\perp = \mathcal{C}(X : Z) \cap \mathcal{C}(X)^\perp \cap \mathcal{C}(Z)^\perp = \{\mathbf{0}\}$.

The matrices $B_i, i = 1, 2, 3$, satisfy

$$R^n = \mathcal{C}(B_1) \boxplus \mathcal{C}(B_2) \boxplus \mathcal{C}(B_3).$$

Thus, a one-to-one transformation $(B_1 : B_2 : B_3)'$ of the model (2.1) yields the equivalent representation through the following three models:

$$B_1'y = B_1'X\beta + B_1'Z\gamma + B_1'\epsilon, \tag{3.1}$$

$$B_2'y = B_2'Z\gamma + B_2'\epsilon, \tag{3.2}$$

$$B_3'y = B_3'\epsilon. \tag{3.3}$$

Since $B_3'y$ is normally distributed and B_3 is orthogonal to $(B_1 : B_2)$, the vector $B_3'y$ is independent of $B_1'y$ as well as $B_2'y$.

In the subsequent, the models (3.1), (3.2), and (3.3) constitute the basis for the derivation of the estimator for $X\beta$.

From construction of B_1 we have $B_1B_1'y = P_X y = \widetilde{X}\beta$, the OLS estimator of $X\beta$, which is functionally independent of the variance components. However, the random effects vector γ is included in both (3.1) and (3.2), and thus for an alternative to the OLS estimator, i.e. a weighted estimator, (3.2) contains essential information about $X\beta$.

Notice $E[B_2'y] = \mathbf{0}$, $B_2'B_2 = (A'ZZ'A)^{-1}$, and $B_2'ZZ'B_2 = I_{l-k}$. Hence $D[B_2'y] = \sigma^2(\rho I_{l-k} + B_2'B_2)$, which implies that the components in $B_2'y$ are correlated and this can cause technical inference problems because if, for example, conditioning $B_1'y$ on $B_2'y$, the inverse of $D[B_2'y]$ is needed and may be difficult to handle.

Now, one can add an extra random term to $B_2'y$, which means adding extra variation, in such a way that the dispersion matrix of the resulting random vector becomes

diagonal, and its inverse is then easy to utilize. This is a key idea of this article. The main remaining issue is how to add the necessary portion of proper variation, since σ_{γ}^2 and σ^2 are both unknown.

The problem can be resolved adopting the ideas of Gallo, Khuri, Öfversten (see the Introduction for references), and others. These authors performed some kind of abstract bootstrapping. Implementing their ideas means to take the important step of adding observable random variables from (3.3) to (3.2) so that eventually a diagonal dispersion matrix is obtained. Let the new random vector \mathbf{u}_2 be defined by

$$\mathbf{u}_2 = \mathbf{B}'_2 \mathbf{y} + (\lambda \mathbf{I}_{l-k} - \mathbf{B}'_2 \mathbf{B}_2)^{1/2} \text{sel}(\mathbf{B}'_3 \mathbf{y}, (l-k)). \quad (3.4)$$

We have introduced here a *selection operator*, $\text{sel}(\mathbf{B}'_3 \mathbf{y}, (l-k))$, which selects $l-k$ independent observations from $\mathbf{B}'_3 \mathbf{y} \sim N_{n-l}(\mathbf{0}, \sigma^2 \mathbf{I})$. This selection may be represented by e.g., an $(l-k) \times (n-l)$ matrix \mathbf{R} , whose rows are the arbitrarily chosen $l-k$ rows of \mathbf{I}_{n-l} . Obviously, $\mathbf{R}\mathbf{R}' = \mathbf{I}_{l-k}$. One possible version of (3.4) is then

$$\mathbf{u}_2 = \mathbf{B}'_2 \mathbf{y} + (\lambda \mathbf{I}_{l-k} - \mathbf{B}'_2 \mathbf{B}_2)^{1/2} \mathbf{R} \mathbf{B}'_3 \mathbf{y}. \quad (3.5)$$

The scalar λ is chosen large enough so that the matrix $\lambda \mathbf{I}_{l-k} - \mathbf{B}'_2 \mathbf{B}_2$ is nonnegative definite and hence the square root in (3.4) exists. There exist several matrix square roots but the different choices do not effect the results in any way.

Theorem 3.1 *Let \mathbf{u}_2 be given by (3.5). Then,*

$$\mathbf{u}_2 \sim N_{(l-k)}(\mathbf{0}, \sigma^2(\rho + \lambda)\mathbf{I}_{l-k}).$$

Proof Since \mathbf{u}_2 is a linear combination of two independently normally distributed random vectors $\mathbf{B}'_2 \mathbf{y}$ and $\mathbf{B}'_3 \mathbf{y}$, it is also normally distributed. Thus, only the first two moments of \mathbf{u}_2 have to be determined. For the expectation, it follows immediately that $E[\mathbf{u}_2] = \mathbf{0}$ because $E[\mathbf{B}'_2 \mathbf{y}] = \mathbf{0}$ and $E[\mathbf{B}'_3 \mathbf{y}] = \mathbf{0}$. Further, direct calculations show that $D[\mathbf{u}_2] = \sigma^2(\rho + \lambda)\mathbf{I}_{l-k}$, which completes the proof. \square

It is clear that if \mathbf{u}_2 is to be used, the parameter λ should be chosen as small as possible (with respect to the existence of the square root) because the dispersion of \mathbf{u}_2 is proportional to λ . Thus, a natural choice for λ is the largest eigenvalue of $\mathbf{B}'_2 \mathbf{B}_2 = (\mathbf{A}' \mathbf{Z} \mathbf{Z}' \mathbf{A})^{-1}$.

Note that all components in \mathbf{u}_2 are mutually independent, which indeed is a very nice property. Moreover, denote $\mathbf{u}_1 = \mathbf{B}'_1 \mathbf{y}$. Then, because of independence between \mathbf{u}_1 and $\mathbf{B}'_3 \mathbf{y}$,

$$\text{cov}[\mathbf{u}_1, \mathbf{u}_2] = \text{cov}[\mathbf{B}'_1 \mathbf{Z} \boldsymbol{\gamma}, \mathbf{B}'_2 \mathbf{Z} \boldsymbol{\gamma}] = \sigma^2 \rho \mathbf{B}'_1 \mathbf{Z} \mathbf{Z}' \mathbf{B}_2.$$

Conditioning \mathbf{u}_1 on \mathbf{u}_2 we obtain

$$\begin{aligned} E[\mathbf{u}_1 | \mathbf{u}_2] &= E[\mathbf{u}_1] + \text{cov}[\mathbf{u}_1, \mathbf{u}_2] D[\mathbf{u}_2]^{-1} \mathbf{u}_2 \\ &= \mathbf{B}'_1 \mathbf{X} \boldsymbol{\beta} + \frac{\rho}{\rho + \lambda} \mathbf{B}'_1 \mathbf{Z} \mathbf{Z}' \mathbf{B}_2 \mathbf{u}_2. \end{aligned}$$

Thus, if the ratio of the variances $\rho = \sigma_y^2/\sigma^2$ is known and because the distribution of \mathbf{u}_2 is independent of $\boldsymbol{\beta}$, the "natural" estimator of $\mathbf{X}\boldsymbol{\beta}$ is given by $\mathbf{B}_1\widehat{\widehat{\mathbf{B}'_1\mathbf{X}\boldsymbol{\beta}}}$, where

$$\widehat{\widehat{\mathbf{B}'_1\mathbf{X}\boldsymbol{\beta}}} = \mathbf{u}_1 - \frac{\rho}{\rho + \lambda} \mathbf{B}'_1\mathbf{Z}\mathbf{Z}'\mathbf{B}_2\mathbf{u}_2. \tag{3.6}$$

In general, the variances or their ratio are not known, hence the ratio

$$\frac{\rho}{\rho + \lambda} = 1 - \frac{\lambda}{\rho + \lambda} \tag{3.7}$$

has to be estimated.

Denote after selection, the remaining variables in (3.3) by

$$\mathbf{u}_3 \sim N_{n-2l+k}(\mathbf{0}, \sigma^2\mathbf{I}_{n-2l+k}). \tag{3.8}$$

Moreover,

$$E[(\mathbf{u}'_2\mathbf{u}_2)^{-1}] = \frac{1}{(l - k - 2)\sigma^2}(\rho + \lambda)^{-1}, \quad l - k > 2, \tag{3.9}$$

since

$$\frac{\mathbf{u}'_2\mathbf{u}_2}{\sigma^2(\rho + \lambda)} \sim \chi^2(l - k),$$

and $E[\mathbf{u}'_3\mathbf{u}_3] = (n - 2l + k)\sigma^2$.

Because \mathbf{u}_2 and \mathbf{u}_3 are independently distributed, it immediately follows that $(\rho + \lambda)^{-1}$ is estimable, i.e.

$$\frac{l - k - 2}{n - 2l + k} E[\mathbf{u}'_3\mathbf{u}_3(\mathbf{u}'_2\mathbf{u}_2)^{-1}] = \frac{1}{\rho + \lambda}, \tag{3.10}$$

and thus also $\mathbf{X}\boldsymbol{\beta}$ can be estimated.

Proposition 3.2 *For the model in (2.1), let \mathbf{u}_i , $i = 1, 2, 3$, and \mathbf{B}_j , $j = 1, 2$, be defined as in the text preceding this proposition. Then, if $n - 2l + k > 0$ and $l - k - 2 > 0$,*

$$\widehat{\mathbf{X}\boldsymbol{\beta}} = \mathbf{B}_1(\mathbf{u}_1 - \mathbf{f}), \tag{3.11}$$

where $\mathbf{f} = (1 - c\mathbf{u}'_3\mathbf{u}_3(\mathbf{u}'_2\mathbf{u}_2)^{-1})\mathbf{B}'_1\mathbf{Z}\mathbf{Z}'\mathbf{B}_2\mathbf{u}_2$, $c = \lambda(l - k - 2)/(n - 2l + k)$, and λ equals the largest eigenvalue of $\mathbf{B}'_2\mathbf{B}_2$.

The proposed estimator for $\mathbf{X}\boldsymbol{\beta}$ is a nonlinear estimator since it is nonlinear in \mathbf{u}_2 and \mathbf{u}_3 . The main objection against this estimator might be that the choice of $l - k$ components from $\mathbf{B}'_3\mathbf{y}$ is arbitrary. Notice that the distribution of \mathbf{u}_2 and \mathbf{u}_3 does not depend on choice of \mathbf{R} in (3.5). Some kind of U -statistic approach may circumvent this type of arbitrariness but this will not be explored in this article.

4 $E[\widehat{X\beta}]$ and $D[\widehat{X\beta}]$

In this section, $E[\widehat{X\beta}]$ and $D[\widehat{X\beta}]$ will be studied, where $\widehat{X\beta}$ was given in Proposition 3.1. The calculations are somewhat lengthy but fairly straightforward. In the following, the next lemma is needed.

Lemma 4.1 *Let u_2 be given by (3.5). Then,*

- (i) $E[u_2(u_2'u_2)^{-1}u_2'] = (l - k)^{-1}I_{l-k};$
- (ii) $E[u_2(u_2'u_2)^{-1}(u_2'u_2)^{-1}u_2'] = (l - k)^{-1}(l - k - 2)^{-1}(\rho + \lambda)^{-1}\frac{1}{\sigma^2}I_{l-k}.$

Proof Let Γ be an arbitrary orthogonal matrix. Since $\Gamma'u_2$ has the same distribution as u_2 , $\Gamma'E[u_2(u_2'u_2)^{-1}u_2']\Gamma = E[u_2(u_2'u_2)^{-1}u_2']$ and $\Gamma'E[u_2(u_2'u_2)^{-1}(u_2'u_2)^{-1}u_2']\Gamma$ equals $E[u_2(u_2'u_2)^{-1}(u_2'u_2)^{-1}u_2']$. Therefore,

$$E[u_2(u_2'u_2)^{-1}u_2'] = c_1I_{l-k},$$

$$E[u_2(u_2'u_2)^{-1}(u_2'u_2)^{-1}u_2'] = c_2I_{l-k},$$

for some constants c_1 and c_2 . Applying the trace-function to these relations yields $c_1 = (l - k)^{-1}$ and

$$c_2 = (l - k)^{-1}E[(u_2'u_2)^{-1}] = (l - k)^{-1}(l - k - 2)^{-1}(\rho + \lambda)^{-1}\frac{1}{\sigma^2},$$

where (3.9) was utilized (see e.g., [Kollo and von Rosen 2005](#); Lemma 2.4.1 for discussion). □

The following moment relations will be used in the subsequent calculations:

$$E[u_1] = B_1'X\beta, \quad E[u_i] = \mathbf{0}, \quad i = 2, 3, \quad E[f] = \mathbf{0},$$

$$E[u_1|u_2] = B_1'X\beta + \rho(\rho + \lambda)^{-1}B_1'ZZ'B_2u_2, \tag{4.1}$$

$$D[u_1] = \sigma^2(\rho B_1'ZZ'B_1 + I_k), \tag{4.2}$$

$$D[u_2] = \sigma^2(\rho + \lambda)I_{l-k}, \tag{4.3}$$

$$E[u_3'u_3] = (n - 2l + k)\sigma^2, \tag{4.4}$$

$$E[(u_3'u_3)^2] = (n - 2l + k)(n - 2(l - 1) + k)(\sigma^2)^2. \tag{4.5}$$

Here we have used $u_3'u_3/\sigma^2 \sim \chi^2(n - 2l + k)$.

Since $E[u_2(u_2'u_2)^{-1}] = \mathbf{0}$, it follows from Proposition 3.2 that for all $\beta \in R^k$,

$$E[\widehat{X\beta}] = E[B_1u_1] = X\beta. \tag{4.6}$$

Furthermore, from Proposition 3.2 it follows that

$$D[\widehat{X\beta}] = B_1(D[u_1] + D[f] - 2\text{cov}[u_1, f])B_1', \tag{4.7}$$

since it can be shown that $\text{cov}[\mathbf{u}_1, \mathbf{f}] = \text{cov}[\mathbf{f}, \mathbf{u}_1]$. The dispersion $D[\mathbf{u}_1]$ was presented in (4.2) and it remains to derive $D[\mathbf{f}]$ and $2\text{cov}[\mathbf{u}_1, \mathbf{f}]$. We start calculating

$$D[\mathbf{f}] = E[(1 - c\mathbf{u}'_3\mathbf{u}_3(\mathbf{u}'_2\mathbf{u}_2)^{-1})^2 \mathbf{B}'_1 \mathbf{Z}\mathbf{Z}' \mathbf{B}_2 \mathbf{u}_2 \mathbf{u}'_2 \mathbf{B}'_2 \mathbf{Z}\mathbf{Z}' \mathbf{B}_1],$$

where as in Proposition 3.2,

$$c = \lambda \frac{l - k - 2}{n - 2l + k}.$$

Hence, we calculate

$$E[\mathbf{B}'_1 \mathbf{Z}\mathbf{Z}' \mathbf{B}_2 \mathbf{u}_2 \mathbf{u}'_2 \mathbf{B}'_2 \mathbf{Z}\mathbf{Z}' \mathbf{B}_1] = \sigma^2(\rho + \lambda) \mathbf{B}'_1 \mathbf{Z}\mathbf{Z}' \mathbf{B}_2 \mathbf{B}'_2 \mathbf{Z}\mathbf{Z}' \mathbf{B}_1,$$

where (4.3) was used;

$$\begin{aligned} & -2E[c\mathbf{u}'_3\mathbf{u}_3(\mathbf{u}'_2\mathbf{u}_2)^{-1} \mathbf{B}'_1 \mathbf{Z}\mathbf{Z}' \mathbf{B}_2 \mathbf{u}_2 \mathbf{u}'_2 \mathbf{B}'_2 \mathbf{Z}\mathbf{Z}' \mathbf{B}_1] \\ & = -2c(n - 2l + k)\sigma^2 E[\mathbf{u}_2(\mathbf{u}'_2\mathbf{u}_2)^{-1} \mathbf{u}'_2] \mathbf{B}'_1 \mathbf{Z}\mathbf{Z}' \mathbf{B}_2 \mathbf{B}'_2 \mathbf{Z}\mathbf{Z}' \mathbf{B}_1 \\ & = -\sigma^2 2c(n - 2l + k)(l - k)^{-1} \mathbf{B}'_1 \mathbf{Z}\mathbf{Z}' \mathbf{B}_2 \mathbf{B}'_2 \mathbf{Z}\mathbf{Z}' \mathbf{B}_1, \end{aligned}$$

where (4.5) and Lemma 4.1 (i) were used;

$$\begin{aligned} & E[c^2(\mathbf{u}'_3\mathbf{u}_3)^2(\mathbf{u}'_2\mathbf{u}_2)^{-2} \mathbf{B}'_1 \mathbf{Z}\mathbf{Z}' \mathbf{B}_2 \mathbf{u}_2 \mathbf{u}'_2 \mathbf{B}'_2 \mathbf{Z}\mathbf{Z}' \mathbf{B}_1] \\ & = c^2(n - 2l + k)(n - 2(l - 1) + k)(\sigma^2)^2 E[\mathbf{u}_2(\mathbf{u}'_2\mathbf{u}_2)^{-1}(\mathbf{u}'_2\mathbf{u}_2)^{-1} \mathbf{u}'_2] \\ & \quad \times \mathbf{B}'_1 \mathbf{Z}\mathbf{Z}' \mathbf{B}_2 \mathbf{B}'_2 \mathbf{Z}\mathbf{Z}' \mathbf{B}_1 \\ & = \sigma^2 c^2 \frac{(n - 2l + k)(n - 2(l - 1) + k)}{(l - k)(l - k - 2)(\rho + \lambda)} \mathbf{B}'_1 \mathbf{Z}\mathbf{Z}' \mathbf{B}_2 \mathbf{B}'_2 \mathbf{Z}\mathbf{Z}' \mathbf{B}_1, \end{aligned}$$

where (4.5) and Lemma 4.1 (ii) were used, which together yield

$$\begin{aligned} D[\mathbf{f}] = \sigma^2 \left(\rho + \lambda - 2c \frac{(n - 2l + k)}{l - k} \right. \\ \left. + c^2 \frac{(n - 2l + k)(n - 2(l - 1) + k)}{(l - k)(l - k - 2)(\rho + \lambda)} \right) \mathbf{B}'_1 \mathbf{Z}\mathbf{Z}' \mathbf{B}_2 \mathbf{B}'_2 \mathbf{Z}\mathbf{Z}' \mathbf{B}_1. \end{aligned} \tag{4.8}$$

Next $\text{cov}[\mathbf{f}, \mathbf{u}_1]$ will be calculated via conditioning $\mathbf{u}_1|\widehat{\mathbf{u}}_2$. Here $E_y[\cdot]$ indicates that expectation is taken with respect to the distribution of \mathbf{y} .

$$\begin{aligned} \text{cov}[\mathbf{f}, \mathbf{u}_1] & = E_{u_2} E_{u_1|u_2} E_{u_3} [(1 - c\mathbf{u}'_3\mathbf{u}_3(\mathbf{u}'_2\mathbf{u}_2)^{-1}) \mathbf{B}'_1 \mathbf{Z}\mathbf{Z}' \mathbf{B}_2 \mathbf{u}_2 \mathbf{u}'_1] \\ & = E_{u_2} [(1 - c(n - 2l + k)\sigma^2(\mathbf{u}'_2\mathbf{u}_2)^{-1}) \mathbf{B}'_1 \mathbf{Z}\mathbf{Z}' \mathbf{B}_2 \mathbf{u}_2 \mathbf{u}'_2 \mathbf{B}'_2 \mathbf{Z}\mathbf{Z}' \mathbf{B}_1] \\ & \quad \times \rho(\rho + \lambda)^{-1} \\ & = \sigma^2 \rho \left(1 - c \frac{(n - 2l + k)}{(l - k)(\rho + \lambda)} \right) \mathbf{B}'_1 \mathbf{Z}\mathbf{Z}' \mathbf{B}_2 \mathbf{B}'_2 \mathbf{Z}\mathbf{Z}' \mathbf{B}_1, \end{aligned} \tag{4.9}$$

where (4.4) and Lemma 4.1 (i) have been applied. Thus, $D[\widehat{\mathbf{X}}\widehat{\boldsymbol{\beta}}]$ is obtained.

Denote the combined constants in $D[f]$ and $\text{cov}[f, \mathbf{u}_1]$ by

$$\begin{aligned} \rho_1 = \rho + \lambda - 2 \frac{l-k-2}{l-k} \lambda + \frac{l-k-2}{l-k} \frac{n-2(l-1)+k}{n-2l+k} \frac{(\lambda)^2}{\rho + \lambda} \\ - 2\rho \left(1 - \frac{(l-k-2)\lambda}{(l-k)(\rho + \lambda)} \right). \end{aligned} \tag{4.10}$$

After substituting (4.8), (4.9), and (4.10) into (4.7), we get

$$D[\widehat{X\beta}] = \sigma^2 P_X + \sigma^2 \rho P_X Z \left[I + \frac{\rho_1}{\rho} Z' B_2 B_2' Z \right] Z' P_X. \tag{4.11}$$

The above obtained results are summarized in the next theorem.

Theorem 4.2 *Let $\widehat{X\beta}$ be given by (3.11) and let $\widetilde{X\beta}$ be the OLS of $X\beta$. Then*

(i) *for all $\beta \in R^k$, $\sigma^2 > 0$, $\rho \geq 0$,*

$$E[\widehat{X\beta}] = X\beta;$$

(ii)

$$D[\widehat{X\beta}] = D[\widetilde{X\beta}] + \sigma^2 \rho_1 P_X Z Z' B_2 B_2' Z Z' P_X,$$

where ρ_1 is given by (4.10).

5 Comparison of $D[\widehat{g'\beta}]$ with the dispersion matrix of the OLS estimator

Let $g'\beta$ be an arbitrary estimable function of β , i.e., $g \in \mathcal{C}(X')$. Equivalently, $g = X'h$ for some h . From (4.11) and Theorem 4.2 we get immediately,

$$D[\widehat{g'\beta}] = D[\widehat{h'X\beta}] = D[\widetilde{g'\beta}] + \sigma^2 \rho_1 h' P_X Z Z' B_2 B_2' Z Z' P_X h.$$

As presented in (4.2), the dispersion matrix of $\widetilde{X\beta} = B_1 u_1 = P_X y$ equals

$$\sigma^2 B_1 (\rho B_1' Z Z' B_1 + I_k) B_1' = \sigma^2 P_X (\rho Z Z' + I_k) P_X,$$

from where it immediately follows,

$$D[\widetilde{g'\beta}] = \sigma^2 h' P_X (\rho Z Z' + I_k) P_X h.$$

Let $d = \frac{l-k-2}{l-k}$. In order to see if the proposed estimator $\widehat{\beta}$ improves the ordinary least squares estimator, according to Theorem 4.2 (ii), the condition under which $\rho_1 < 0$, i.e.

$$\rho + \lambda - 2d\lambda + d \frac{n-2(l-1)+k}{n-2l+k} \frac{(\lambda)^2}{\rho + \lambda} - 2 \left(\rho - d \frac{\rho\lambda}{\rho + \lambda} \right) < 0$$

Table 1 Configuration settings and calculated λ for the simulation study

Configuration	σ^2	λ	n	l	n_1	n_2	n_3	n_4	n_5	n_6	n_7	n_8	n_9	n_{10}
1	1	7.175	30	3	13	11	6	-	-	-	-	-	-	-
2	1	2.089	38	10	3	3	3	3	4	7	4	2	5	4

In all cases $\beta = 10$

has to be studied. This expression is equivalent to

$$\rho > \lambda + d \left(\frac{n-2(l-1)+k}{n-2l+k} - 2 \right) \frac{\lambda^2}{\rho+\lambda}, \tag{5.1}$$

where it has been used that

$$\frac{\rho}{\rho+\lambda} = 1 - \frac{\lambda}{\rho+\lambda}.$$

Theorem 5.1 Let $\widehat{X\beta}$ be given by (3.11). Then for every estimable function $g'\beta$, $D[\widehat{g'\beta}] - D[g'\beta] > 0$ if and only if

$$\rho > \lambda \left(1 - \left(1 - \frac{2}{n-2l+k} \right) \left(1 - \frac{2}{l-k} \right) \right)^{\frac{1}{2}}.$$

Proof Manipulating (5.1) yields that $\widehat{X\beta}$ has a smaller dispersion matrix if and only if

$$\rho - \lambda + \frac{\lambda^2}{\rho+\lambda} \frac{n-2l+k-2}{n-2l+k} \frac{l-k-2}{l-k} > 0.$$

The statement of the theorem is the solution to this inequality. □

From Theorem 5.1 it follows that the inequality $\rho > \lambda$ is a simple criterion for deciding if $\widehat{X\beta}$ should be used instead of the OLS estimator, although ρ has to be estimated. Here we can use as variance estimator $\widehat{\sigma}^2 = (n-2l+k)^{-1} \mathbf{u}'_3 \mathbf{u}_3$. However, if we use the estimator $\widehat{\rho} = \frac{n-2(l+1)+k}{l-k} \mathbf{u}'_2 \mathbf{u}_2 (\mathbf{u}'_3 \mathbf{u}_3)^{-1} - \lambda$, which is motivated by (3.8), we observe that if $\rho \ll \lambda$, $\widehat{\rho}$ can take negative values. Similarly, if we use the estimator $\widehat{\sigma}_y^2 = (l-k)^{-1} \mathbf{u}'_2 \mathbf{u}_2 - \lambda \widehat{\sigma}^2$, which is also motivated by Theorem 3.1 and (3.8), we can observe that if $\sigma_y^2 \ll \sigma^2$, $\widehat{\sigma}_y^2$ can take negative values, indicating that the OLS estimator is preferable.

6 Example: unbalanced one-way random model

A special case of model (2.1) is the one-way random effects model

$$y_{ij} = \beta + \gamma_i + \epsilon_{ij}, \tag{6.1}$$

Table 2 Empirical estimates, and the theoretical variances $D[\tilde{\beta}]$ and $D[\hat{\beta}]$ for Configuration 1

$\rho = \frac{\sigma_\gamma^2}{\sigma^2}$	$\tilde{\beta}$ (OLS)	$\hat{\beta}$	$\hat{\sigma}^2$	$\hat{\sigma}_\gamma^2$	$D[\tilde{\beta}]$	$D[\hat{\beta}]$	$\widehat{MSE}(\tilde{\beta})$	$\widehat{MSE}(\hat{\beta})$
0.10	9.996	9.995	0.997	0.036	0.070	0.274	0.069	0.269
0.25	9.995	9.996	0.999	0.312	0.124	0.324	0.121	0.326
0.50	9.997	9.996	1.001	0.700	0.214	0.407	0.214	0.419
1.00	9.910	9.999	1.003	1.024	0.396	0.574	0.394	0.567
1.50	9.996	9.993	1.001	1.486	0.577	0.741	0.577	0.739
2.00	9.994	9.999	0.999	2.123	0.758	0.907	0.763	0.923
2.50	10.017	10.009	1.007	2.361	0.939	1.074	0.938	1.066
5	10.007	10.008	0.997	4.873	1.844	1.907	1.866	1.933
10	9.996	9.991	0.998	10.080	3.656	3.574	3.541	3.509
20	10.054	10.039	0.996	19.954	7.278	6.907	7.334	6.912

Bold values indicate the instances when the newly suggested estimator has smaller variance than the ordinary least squares estimator

Table 3 Empirical estimates, and the theoretical variances $D[\tilde{\beta}]$ and $D[\hat{\beta}]$ for Configuration 2

$\rho = \frac{\sigma_\gamma^2}{\sigma^2}$	$\tilde{\beta}$ (OLS)	$\hat{\beta}$	$\hat{\sigma}^2$	$\hat{\sigma}_\gamma^2$	$D[\tilde{\beta}]$	$D[\hat{\beta}]$	$\widehat{MSE}(\tilde{\beta})$	$\widehat{MSE}(\hat{\beta})$
0.10	9.999	10.000	0.999	0.108	0.037	0.045	0.037	0.044
0.25	10.000	10.000	0.994	0.262	0.054	0.061	0.056	0.062
0.50	10.007	10.007	0.992	0.529	0.082	0.087	0.082	0.087
1.00	9.996	9.993	1.004	0.986	0.138	0.140	0.140	0.142
1.50	10.003	10.002	0.999	1.475	0.195	0.191	0.196	0.194
2.00	9.999	9.998	0.999	1.964	0.251	0.243	0.257	0.252
2.50	9.992	9.993	1.001	2.486	0.307	0.294	0.307	0.299
5	9.991	9.988	1.001	4.999	0.587	0.547	0.604	0.569
10	9.996	9.994	1.007	9.961	1.148	1.049	1.155	1.076
20	10.017	10.015	1.002	20.015	2.270	2.050	2.200	2.033

Bold values indicate the instances when the newly suggested estimator has smaller variance than the ordinary least squares estimator

where $i = 1, \dots, l, j = 1, \dots, n_i$, and $n = \sum_{i=1}^l n_i$. The expectation $\beta \in R$ is our parameter of interest. All distributional assumptions of model (2.1) are assumed to hold. Here $k = 1$, the matrix X is an n -dimensional column vector of ones, $X = \mathbf{1}_n$, and the matrix Z is a block matrix of column vectors of ones of length n_i , $Z = \text{Diag}\{\mathbf{1}_{n_i}\}$. The OLS estimator of β is $\bar{y} = \mathbf{u}_1$, the average of all n observations. It is well known that the maximum likelihood estimator (MLE) of μ , in case of an unbalanced model, has to be obtained iteratively. Note that $C(X) \subseteq C(Z)$ and thus $l = \text{rank}(Z)$.

We illustrate our procedure using a small simulation study for model (6.1). A broad range for the ratio of the variance components ρ is considered. The number of levels l of γ is chosen as small as $l = 3$ with relatively higher number of observations n_i

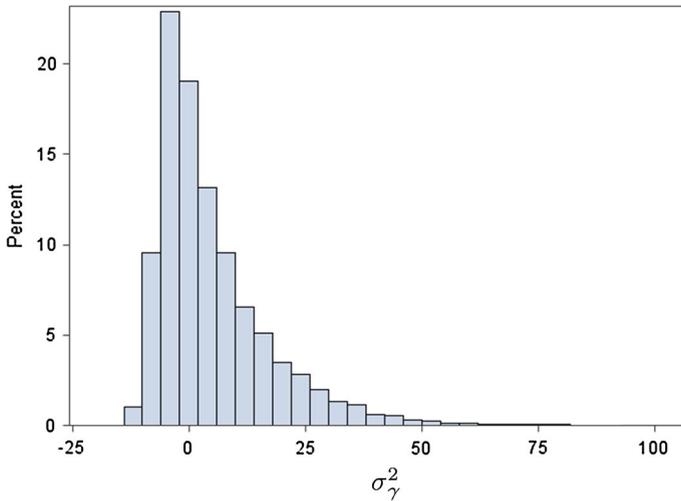


Fig. 1 Histogram from 10,000 estimates of $\widehat{\sigma}_\gamma^2$ for Configuration 2, for $\rho = 5$, illustrating positive probability of negative estimates of σ_γ^2

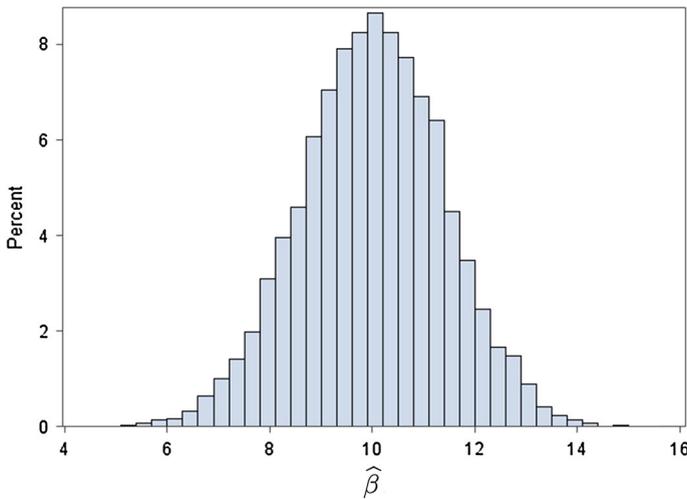


Fig. 2 Histogram from 10,000 estimates of $\widehat{\beta}$ for Configuration 2, for $\rho = 5$, illustrating the symmetry of the distribution of $\widehat{\beta}$

per level, and a relatively higher $l = 10$, with smaller numbers n_i . Because of the proportionality, we considered $\sigma^2 = 1$ only.

All setting configurations are presented in Table 1. For each configuration, 10,000 simulations were carried out. The OLS estimates as well as $\widehat{\beta}$ estimates are presented as averages of 10,000 observed estimates. In addition, the estimates of σ^2 and of σ_γ^2 are presented. The observed MSE of the OLS estimators as well as of the $\widehat{\beta}$ s are also included here. The results are presented in Tables 2 and 3.

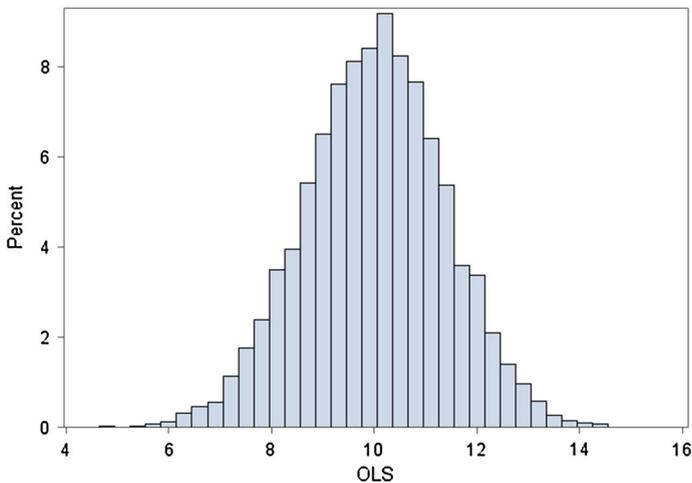


Fig. 3 Histogram from 10,000 estimates of the OLS estimator for Configuration 2, for $\rho = 5$

In Sect. 5, it was suggested that the new proposed estimator has a smaller dispersion if $\rho > \lambda = 7.175$ which is in complete agreement with Table 2 where for $\rho = 10$ or $\rho = 20$ simulations indicate the new estimator has a smaller dispersion than the OLS estimator, but not for $\rho = 5$. The results for Configuration 2 are presented in Table 3. Now using Table 3 and Theorem 5.1, it follows that ρ should be larger than 1.97 if $\hat{\beta}$ is to be applied instead of the OLS estimator. This strategy is supported by results of simulations presented in Table 3. The tables indicate that even a smaller ρ can be used but one has to remember that estimated variances and MSEs are applied, in particular that $\hat{\sigma}_y$ can become negative (see Figure 1). However, with confidence we can state that the new estimator is better than the least squares estimator in certain regions of the parameter space (described through ρ) as it is shown in Theorem 5.1 and simulations. In addition, we would like to point out that in spite of $\hat{\beta}$ being a nonlinear estimator, it is an unbiased estimator as is shown in (4.6), and as we observed in our simulations, its distribution seems to be symmetric around its expectation. For illustration see histograms of $\hat{\beta}$ and of the OLS estimators in Figure 2 and Figure 3, respectively.

Acknowledgements The research has been supported by The Swedish Foundation for Humanities and Social sciences (P14-0641:1).

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Al Sarraj R, von Rosen D (2009) Improving Henderson's method 3 approach when estimating variance components in a two-way mixed linear model. In: Schipp B, Krämer W (eds) Statistical inference, econometric analysis and matrix algebra. Physica-Verlag, Heidelberg, pp 125–142

- Christensen R (1996) Exact tests for variance components. *Biometrics* 52:309–314
- Gallo J, Khuri AI (1990) Exact tests for the random and fixed effects in an unbalanced mixed two-way cross-classification model. *Biometrics* 46:1087–1095
- Güven B (2012) Approximate tests in unbalanced two-way random models without interaction. *Stat Pap* 53:753–766
- Henderson CR (1953) Estimation of variance and covariance components. *Biometrics* 9:226–252
- Khuri AI (1986) Exact tests for the comparison of correlated response models with an unknown dispersion matrix. *Technometrics* 28:347–357
- Khuri AI (1990) Exact tests for random models with unequal cell frequencies in the last stage. *J Stat Plan Inference* 24:177–193
- Khuri AI, Littell RC (1987) Exact tests for the main effects variance components in an unbalanced random two-way model. *Biometrics* 43:545–560
- Kollo T, von Rosen D (2005) *Advanced multivariate statistics with matrices. Mathematics and its applications*, vol 579. Springer, Dordrecht
- Öfversten J (1993) Exact tests for variance components in unbalanced mixed linear models. *Biometrics* 49:45–57
- Öfversten J (1995) Estimation in mixed models via layer triangular transformation. *Comput Stat Data Anal* 20:657–667